nhanes_metals

Nicole Comfort

11/26/2019

First, check whether you have the necessary packages installed and if not, install them.

Background:

The National Health and Nutrition Examination Survey (NHANES) is a nationally-representative survey of the general, noninstutionalized US civilian population conducted by the National Center for Health Statistics. NHANES is a multi-stage, nationally representative sample in two-year cycles; participants are chosen to represent the entire general US. Certain age groups and racial/ethnic groups are intentionally oversampled to ensure adequate power to analyze these populations.

Participants undergo thorough examination including demographic questionnaire, a 24-hr dietary recall, a clinical examination, and a laboratory examination. All NHANES protocols were approved by the National Center for Health Statistics institutional review board, and all participants gave written informed consent.

For more information on NHANES, view the Centers for Disease Control and Prevention (CDC) website: https://www.cdc.gov/nchs/nhanes/index.htm

Note:

NHANES does not measure every environmental chemical in every participant. For some environmental chemicals, these measurements are only done on a subset of participant samples and there are special sub-weights that need to be used for analyzing these chemicals. As a general rule, you should use the weight for the smallest of the sampling subset you are analyzing in NHANES.

Here, we will explore the levels of lead, cadmium, mercury, and manganese measured in blood of males versus females to examine if there is a sex difference in body burden of these chemicals.

We will conduct this analysis in one two-year survey cycle (2015/2016) to start, *since combining weights across survey cycles is very complex.* See http://www.cdc.gov/nchs/tutorials/NHANES/SurveyDesign/Weighting/intro.htm for more information.

Download NHANES Data and Clean

The data files must first be downloaded. They can be found and downloaded at the NHANES website: https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2015

They must first be downloaded.

```
# Load data

# Demographic files for 2015-2016
demo1516 <- read.xport("DEMO_I.XPT")
## Data documentation file: https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DEMO_I.htm

# Laboratory blood Pb, Cd, Hg, Mn files for 2015-2016
cd1516 <- read.xport("PBCD_I.XPT")
## Data documentation file: https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/PBCD_I.htm
```

```r
### Eligible Sample: All examined participants aged 1-11 years old, and a one-half sample from particip

# Inorganic, Ethyl, Methyl Mercury for 2015-2016
hg1516 <- read.xport("IHGEM_I.XPT")
## Data documentation file: https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/IHGEM_I.htm
### Eligible Sample: All examined participants aged 1-11 years old, and a one-half sample from particip
```

```r
# Merge data to create one dataframe per year; then keep only the necessary variables:
nh1516 <- merge(demo1516, cd1516, by = "SEQN", all = TRUE)
nh1516 <- merge(nh1516, hg1516, by = c("SEQN", "WTSH2YR"), all = TRUE) # exclude those who have this va

# double check that there aren't repeat columns renamed [col_name].x or [col_name].y

## Here is a list of variables we will need for our analysis:
# SEQN, SDDSRVYR, RIAGENDR, RIDAGEYR, RIDRETH1, WTMEC2YR, SDMVPSU, SDMVSTRA, WTSH2YR
# LBXBPB (blood lead), LBXBCD (blood cadmium),
# LBXTHG (blood total mercury), LBXIHG (blood inorganic mercury),
# LBXBSE (blood selenium), LBXBMN (blood manganese)

## any more?  (11/26/19 7:17pm)

# Retain only necessary variables:
nh1516_sub <-
  nh1516 %>%
  select(SEQN, SDDSRVYR, # unique identifier and ?
         RIAGENDR, RIDAGEYR, RIDRETH1, # gender, age, race/ethnicity
         WTMEC2YR, SDMVPSU, SDMVSTRA, WTSH2YR, # weighting info
         LBXBPB, LBXBCD, LBXTHG, LBXIHG, LBXBMN) # metals (exclude selenium for now)

head(nh1516_sub)
```

```
##    SEQN SDDSRVYR RIAGENDR RIDAGEYR RIDRETH1  WTMEC2YR SDMVPSU SDMVSTRA
## 1 83732        9        1       62        3 135629.51       1      125
## 2 83733        9        1       53        3  25282.43       1      125
## 3 83734        9        1       78        3  12575.84       1      131
## 4 83735        9        2       56        3 102078.63       1      131
## 5 83736        9        2       42        4  18234.74       2      126
## 6 83737        9        2       72        1  10878.68       1      128
##     WTSH2YR LBXBPB LBXBCD LBXTHG LBXIHG LBXBMN
## 1 283573.31   0.88   0.20   0.47   0.19   7.83
## 2  50411.33   2.60   3.53   3.08   0.19   7.18
## 3  26293.48   1.85   0.43   0.70   0.19   6.18
## 4        NA     NA     NA     NA     NA     NA
## 5        NA     NA     NA     NA     NA     NA
## 6        NA     NA     NA     NA     NA     NA
```

```r
# clean / rename variable names
nh1516_sub <-
  nh1516_sub %>%
  janitor::clean_names() %>%
  rename(sex = riagendr,
         age = ridageyr,
         race = ridreth1) %>%
  mutate(lbxohg = lbxthg - lbxihg) # create new variable for organic mercury (total Hg - inorganic Hg)
```

```
# why are some organic mercury values negative??

glimpse(nh1516_sub)
```

```
## Observations: 9,971
## Variables: 15
## $ seqn     <dbl> 83732, 83733, 83734, 83735, 83736, 83737, 83738, 8373...
## $ sddsrvyr <dbl> 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9,...
## $ sex      <dbl> 1, 1, 1, 2, 2, 2, 2, 1, 1, 1, 2, 1, 1, 2, 2, 1, 1, 2,...
## $ age      <dbl> 62, 53, 78, 56, 42, 72, 11, 4, 1, 22, 32, 18, 56, 15,...
## $ race     <dbl> 3, 3, 3, 3, 4, 1, 1, 3, 2, 4, 1, 5, 4, 3, 5, 3, 4, 3,...
## $ wtmec2yr <dbl> 135629.507, 25282.426, 12575.839, 102078.635, 18234.7...
## $ sdmvpsu  <dbl> 1, 1, 1, 1, 2, 1, 1, 2, 1, 2, 1, 2, 2, 2, 2, 1, 2, 2,...
## $ sdmvstra <dbl> 125, 125, 131, 131, 126, 128, 120, 124, 119, 128, 125...
## $ wtsh2yr  <dbl> 283573.312, 50411.334, 26293.484, NA, NA, NA, 9860.62...
## $ lbxbpb   <dbl> 0.88, 2.60, 1.85, NA, NA, NA, 0.49, 0.29, NA, 0.72, 0...
## $ lbxbcd   <dbl> 0.20, 3.53, 0.43, NA, NA, NA, 0.14, 0.07, NA, 0.20, 0...
## $ lbxthg   <dbl> 0.47, 3.08, 0.70, NA, NA, NA, 0.46, 0.40, NA, 1.38, 1...
## $ lbxihg   <dbl> 0.19, 0.19, 0.19, NA, NA, NA, 0.65, 0.19, NA, 0.19, 0...
## $ lbxbmn   <dbl> 7.83, 7.18, 6.18, NA, NA, NA, 12.57, 10.06, NA, 7.27,...
## $ lbxohg   <dbl> 0.28, 2.89, 0.51, NA, NA, NA, -0.19, 0.21, NA, 1.19, ...
```

```
# Label variable values
nh1516_sub <-
  nh1516_sub %>%
  set_value_labels(
    sex = c("Male" = 1, "Female" = 2),
    race = c("Mexican American" = 1, "Other Hispanic" = 2, "NH White" = 3, "NH Black" = 4, "Other inclu

glimpse(nh1516_sub)
```

```
## Observations: 9,971
## Variables: 15
## $ seqn     <dbl> 83732, 83733, 83734, 83735, 83736, 83737, 83738, 8373...
## $ sddsrvyr <dbl> 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9,...
## $ sex      <dbl+lbl> 1, 1, 1, 2, 2, 2, 2, 1, 1, 1, 2, 1, 1, 2, 2, 1, 1...
## $ age      <dbl> 62, 53, 78, 56, 42, 72, 11, 4, 1, 22, 32, 18, 56, 15,...
## $ race     <dbl+lbl> 3, 3, 3, 3, 4, 1, 1, 3, 2, 4, 1, 5, 4, 3, 5, 3, 4...
## $ wtmec2yr <dbl> 135629.507, 25282.426, 12575.839, 102078.635, 18234.7...
## $ sdmvpsu  <dbl> 1, 1, 1, 1, 2, 1, 1, 2, 1, 2, 1, 2, 2, 2, 2, 1, 2, 2,...
## $ sdmvstra <dbl> 125, 125, 131, 131, 126, 128, 120, 124, 119, 128, 125...
## $ wtsh2yr  <dbl> 283573.312, 50411.334, 26293.484, NA, NA, NA, 9860.62...
## $ lbxbpb   <dbl> 0.88, 2.60, 1.85, NA, NA, NA, 0.49, 0.29, NA, 0.72, 0...
## $ lbxbcd   <dbl> 0.20, 3.53, 0.43, NA, NA, NA, 0.14, 0.07, NA, 0.20, 0...
## $ lbxthg   <dbl> 0.47, 3.08, 0.70, NA, NA, NA, 0.46, 0.40, NA, 1.38, 1...
## $ lbxihg   <dbl> 0.19, 0.19, 0.19, NA, NA, NA, 0.65, 0.19, NA, 0.19, 0...
## $ lbxbmn   <dbl> 7.83, 7.18, 6.18, NA, NA, NA, 12.57, 10.06, NA, 7.27,...
## $ lbxohg   <dbl> 0.28, 2.89, 0.51, NA, NA, NA, -0.19, 0.21, NA, 1.19, ...
```

```
# Treat any labelled variables (e.g. sex) as factors
nh1516_sub <-
  nh1516_sub %>%
  mutate_if(is.labelled, to_factor)
```

Let's evaluate the missing data. Blood metals are measured in all participants aged 1-11 years old, and

additionally in only a one-half subsample of participants 12 years and older. Let's find how many people are missing our main variables of interest: blood lead, cadmium, total and inorganic mercury, and manganese.

```
# assess missingness
# for Pb
nh1516_sub %>%
  count(is.na(lbxbpb))
```

```
## # A tibble: 2 x 2
##   `is.na(lbxbpb)`     n
##   <lgl>           <int>
## 1 FALSE            4988
## 2 TRUE             4983
```

```
# for Cd
nh1516_sub %>%
  count(is.na(lbxbcd))
```

```
## # A tibble: 2 x 2
##   `is.na(lbxbcd)`     n
##   <lgl>           <int>
## 1 FALSE            4988
## 2 TRUE             4983
```

```
# for total Hg
nh1516_sub %>%
  count(is.na(lbxthg))
```

```
## # A tibble: 2 x 2
##   `is.na(lbxthg)`     n
##   <lgl>           <int>
## 1 FALSE            4988
## 2 TRUE             4983
```

```
# for inorganic Hg
nh1516_sub %>%
  count(is.na(lbxihg))
```

```
## # A tibble: 2 x 2
##   `is.na(lbxihg)`     n
##   <lgl>           <int>
## 1 FALSE            4938
## 2 TRUE             5033
```

```
# for Mn
nh1516_sub %>%
  count(is.na(lbxbmn))
```

```
## # A tibble: 2 x 2
##   `is.na(lbxbmn)`     n
##   <lgl>           <int>
## 1 FALSE            4987
## 2 TRUE             4984
```

```
# remove those with NAs

## should I create a separate dataset for each?? or only retain those that are COMPLETE CASES?? (Ask An
# for now (11/27/19 10:30am), I will retain only complete cases
```

```r
# drop rows with missing values
nh1516_sub <-
  nh1516_sub %>%
  drop_na()
```

After retaining only complete cases with no missing values, we are left with observations.

Sample Weighting

Note that NHANES is designed to sample larger numbers of certain subgroups of particular public health interest (e.g. low income Americans, African Americans). Oversampling is done to increase the reliability and precision of estimates of health status indicators for these population subgroups. To account for this sampling design and these design features in our analysis, it is critical that we apply sample weights.

Continue data cleaning by applying appropriate sample weights:

```r
### NOTE: the correct weights to use for this analysis of blood metal data are.... WTSH2YR (double chec

## from NHANES:

############################################################################
# The appropriate sample weights are provided in the variable WTSH2YR in this data file for all partici

# The analytes included in this dataset were measured for all examined participants aged 1-11 years, an

# For participants 12 years and older, special sample weights were created for the subsample. These spe
############################################################################

# you have to choose the survey weights that correspond to the smallest subsample that is relevant for
# I think that these metals (Pb, Cd, Hg, Mn) were measured in everybody, check teh data documentation a
```

Survey Weights

```
## Stratified 1 - level Cluster Sampling design (with replacement)
## With (30) clusters.
## svydesign(ids = ~sdmvpsu, strata = ~sdmvstra, weights = ~wtsh2yr,
##     nest = TRUE, data = nh1516_sub)
## Probabilities:
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 2.001e-06 1.547e-05 2.693e-05 3.914e-05 5.924e-05 1.882e-04
## Stratum Sizes:
##            119 120 121 122 123 124 125 126 127 128 129 130 131 132 133
## obs        205 359 348 275 307 288 318 385 327 357 350 387 407 390 234
## design.PSU   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2
## actual.PSU   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2
## Data variables:
##  [1] "seqn"     "sddsrvyr" "sex"      "age"      "race"     "wtmec2yr"
##  [7] "sdmvpsu"  "sdmvstra" "wtsh2yr"  "lbxbpb"   "lbxbcd"   "lbxthg"
## [13] "lbxihg"   "lbxbmn"   "lbxohg"
```

Descriptive Statistics

Find the mean levels and SE of Pb, Cd, Hg (inorganic and organic), and Mn in the US population (accounting for survey weights and design).

```
# useful survey functions:

## svymean: Computes means and SEs for survey data
svymean(~lbxbpb, # continuous variable(s)
        nhanes.svy) # the "svydesign" object
```

```
##          mean     SE
## lbxbpb 1.0794 0.0322
```

```
# Example, we compute the means of multiple variables:
svymean(~lbxbpb+lbxbcd+lbxthg+lbxihg+lbxohg+lbxbmn,
        nhanes.svy)
```

```
##            mean     SE
## lbxbpb  1.07937 0.0322
## lbxbcd  0.36494 0.0108
## lbxthg  1.19763 0.0779
## lbxihg  0.24118 0.0057
## lbxohg  0.95645 0.0803
## lbxbmn 10.14609 0.1126
```

```
# put results into a table
mean_metals_table <-
  svymean(~lbxbpb+lbxbcd+lbxthg+lbxihg+lbxohg+lbxbmn, nhanes.svy) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = F)
# note that lbxbpb is in the units ug/dL whereas the others are in ug/L

#################################################################
## By Sex

## svyby: Survey stats on subsets of data (or, survey stats across levels of factors)
svyby(~lbxbpb+lbxbcd+lbxthg+lbxihg+lbxohg+lbxbmn, # continuous variable
      by = ~sex, # groups
      nhanes.svy, # the "svydesign" object
      svymean)  # the function you want %>%
```

```
##           sex    lbxbpb    lbxbcd   lbxthg    lbxihg    lbxohg    lbxbmn
## Male     Male 1.2157178 0.3420003 1.222163 0.2315373 0.9906259  9.472531
## Female Female 0.9496064 0.3867703 1.174287 0.2503628 0.9239243 10.787150
##         se.lbxbpb  se.lbxbcd  se.lbxthg   se.lbxihg  se.lbxohg se.lbxbmn
## Male   0.04206694 0.01457146 0.07485000 0.005234537 0.07640435 0.1226099
## Female 0.03830881 0.01431977 0.09065983 0.008886026 0.09375362 0.1229766
```

```
# put into a table
sex_mean_metals_table <-
  svyby(~lbxbpb+lbxbcd+lbxthg+lbxihg+lbxohg+lbxbmn, # continuous variable
      by = ~sex, # groups
      nhanes.svy, # the "svydesign" object
      svymean) %>%  # the function you want %>%
  kable() %>%
```

```
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = F)

#################################################################

# Other useful functions:

## svyhist: Survey-weighted histograms

## svyquantile: Computes quantiles for survey data (you can choose the quantile values)
svyquantile(~lbxbpb, # continuous variable
            nhanes.svy, # the "svydesign" object
            c(0.10, 0.90), # list of percentiles
            ci = F)
```

```
##         0.1      0.9
## lbxbpb 0.35 2.137056
```

```
## svychisq: Contingency table and chi-squared tests for survey data; weighted cross tabulations
```

The following are the estimated means and SEs of blood levels of lead, cadmium, mercury (total, inorganic, and organic), and manganese in the US population:

Broken down by sex:

Note that, as expected, males have higher mean Pb levels. It also appears that females have higher Mn levels. Are these differences significant? Let's conduct t tests.

```
## svyttest: t tests

# Lead (Pb)
#####################################

# To determine if this difference is statistically significant, I do a t-test because I am comparing th
tt_pb <-
  svyttest(lbxbpb~sex,
           nhanes.svy)

tt_pb
```

```
##
##  Design-based t-test
##
## data:  lbxbpb ~ sex
## t = -5.5163, df = 14, p-value = 7.598e-05
## alternative hypothesis: true difference in mean is not equal to 0
## 95 percent confidence interval:
##  -0.3606616 -0.1715612
## sample estimates:
## difference in mean
##         -0.2661114
```

```
## What if I want a 90% CI instead of a 95% CI?
# confint(tt, level=0.8)

# Manganese (Mn)
#####################################
```

```r
tt_mn <-
  svyttest(lbxbmn~sex,
           nhanes.svy)

tt_mn
```

```
##
##  Design-based t-test
##
## data:  lbxbmn ~ sex
## t = 13.563, df = 14, p-value = 1.915e-09
## alternative hypothesis: true difference in mean is not equal to 0
## 95 percent confidence interval:
##  1.124641 1.504598
## sample estimates:
## difference in mean
##           1.314619
```

The design-based t-tests indicate that the difference in both lead and manganese levels across males and females is significant.

Differences across age groups

The exposures and body burdens of these various metals often change over the life course, from childhood through elderly years. Do sex differences exist at certain windows across the life course that are otherwise masked when looking at mean metal levels averaged across all ages?

To explore this, we will create a new categorical variable, "life_stage," that indicates whether an individual is in his or her child, adolescent, adult, or senior years.

Question: Accounting for survey design, what are the mean blood metal levels of 2015/2016 US males and females across childhood, adolescence, adulthood, and senior years?

```r
## create new life stage variable
nh1516_sub <-
  nh1516_sub %>%
  mutate(life_stage =
           ifelse(age < 12, "Child",
           ifelse(age %in% 12:21, "Adolescent",
           ifelse(age %in% 22:65, "Adult",
                  "Elderly")))) %>%
  mutate(life_stage = as.factor(life_stage))

# relevel the factors to be in chronological order of life stages
nh1516_sub$life_stage <- fct_relevel(nh1516_sub$life_stage, c("Child", "Adolescent", "Adult", "Elderly"

# Re-set the survey design
# NOTE: We always have to re-set the survey design after we create any new variable.
nhanes.svy <- svydesign(ids = ~sdmvpsu,
                        strata = ~sdmvstra,
                        weights = ~wtsh2yr,
                        nest = TRUE,
                        data = nh1516_sub)
```

```
## Calculate means across sex and life stages
svyby(~lbxbpb+lbxbcd+lbxthg+lbxihg+lbxohg+lbxbmn, # continuous variable
      by = ~sex+life_stage, # groups
      nhanes.svy, # the "svydesign" object
      svymean)  # the function you want
```

```
##                       sex life_stage    lbxbpb     lbxbcd    lbxthg
## Male.Child           Male      Child 0.8501239 0.09863372 0.4459588
## Female.Child       Female      Child 0.7786317 0.10267393 0.4388593
## Male.Adolescent      Male Adolescent 0.6394596 0.17705843 0.7431200
## Female.Adolescent Female Adolescent 0.4738220 0.18132393 0.4868935
## Male.Adult           Male      Adult 1.2766305 0.40530379 1.4361154
## Female.Adult       Female      Adult 0.9328629 0.46293280 1.4192908
## Male.Elderly         Male    Elderly 1.8764311 0.42088906 1.3584639
## Female.Elderly     Female    Elderly 1.4773593 0.43594691 1.2410247
##                       lbxihg    lbxohg    lbxbmn  se.lbxbpb   se.lbxbcd
## Male.Child         0.2106070 0.2353518 10.632091 0.05914193 0.002326622
## Female.Child       0.2089638 0.2298955 11.389825 0.04839417 0.002702114
## Male.Adolescent    0.2266632 0.5164568 10.286462 0.01905087 0.015844596
## Female.Adolescent  0.2406871 0.2462064 11.772936 0.02776808 0.018057835
## Male.Adult         0.2314264 1.2046890  9.175066 0.05565521 0.023564370
## Female.Adult       0.2584671 1.1608238 10.859249 0.04168051 0.018552600
## Male.Elderly       0.2573128 1.1011512  9.043854 0.07559327 0.029916590
## Female.Elderly     0.2533815 0.9876432  9.399506 0.06438022 0.017896607
##                      se.lbxthg   se.lbxihg  se.lbxohg se.lbxbmn
## Male.Child         0.02892840 0.005262277 0.02985143 0.1958392
## Female.Child       0.03181080 0.007083769 0.03110242 0.2087941
## Male.Adolescent    0.12128072 0.013109366 0.12076355 0.3144730
## Female.Adolescent  0.03235831 0.029758228 0.05472757 0.4017955
## Male.Adult         0.09984121 0.007160509 0.10179784 0.1596139
## Female.Adult       0.11172990 0.010557135 0.11565427 0.1047459
## Male.Elderly       0.10608952 0.015512391 0.10670209 0.2692876
## Female.Elderly     0.16860522 0.017603869 0.16931400 0.2639640
```

```
life_means_table <-
  svyby(~lbxbpb+lbxbcd+lbxthg+lbxihg+lbxohg+lbxbmn,
      by = ~sex+life_stage,
      nhanes.svy,
      svymean) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = F)
```

Across sex and age groups:

```
# first, create a new dataset from the life_means_table data
# (did manually by copying and pasting the data into a csv)
# import data
life_means_data <-
  read_csv("life_means.csv") %>%
  mutate(life_stage = as.factor(life_stage))
```

```
## Parsed with column specification:
## cols(
##   Sex = col_character(),
```

```
##   life_stage = col_character(),
##   Pb = col_double(),
##   Cd = col_double(),
##   `Total Hg` = col_double(),
##   `Inorganic Hg` = col_double(),
##   `Organic Hg` = col_double(),
##   Mn = col_double(),
##   se.lbxbpb = col_double(),
##   se.lbxbcd = col_double(),
##   se.lbxthg = col_double(),
##   se.lbxihg = col_character(),
##   se.lbxohg = col_double(),
##   se.lbxbmn = col_double()
## )
```

```r
life_means_data$life_stage <-
  factor(life_means_data$life_stage, levels = c("Child", "Adolescent", "Adult", "Elderly"))

# Lead
pb_plot <-
life_means_data %>%
  na.omit() %>%
  ggplot(aes(x = life_stage,
             y = Pb,  # change for different metals
             fill = Sex)) +
  geom_col(position = "dodge") +
  geom_errorbar(aes(ymin = Pb - se.lbxbpb, # change for different metalas
                    ymax = Pb + se.lbxbpb)) +
  labs(title="Levels of Blood Pb in Males vs Females Across Life Stages", # edit labels for diff graphs
       x = "Life Stage",
       y = "Average Blood Pb (ug/dL)",
       caption = "NHANES 2015/2016 data")
### Need to figure out how to dodge the error bar

# Cadmium
cd_plot <-
life_means_data %>%
  na.omit() %>%
  ggplot(aes(x = life_stage,
             y = Cd,
             fill = Sex)) +
  geom_col(position = "dodge") +
  geom_errorbar(aes(ymin = Cd - se.lbxbcd,
                    ymax = Cd + se.lbxbcd)) +
  labs(title="Levels of Blood Cd in Males vs Females Across Life Stages",
       x = "Life Stage",
       y = "Average Blood Pb (ug/L)",
       caption = "NHANES 2015/2016 data")

# Manganese
mn_plot <-
life_means_data %>%
  na.omit() %>%
  ggplot(aes(x = life_stage,
```

```
             y = Mn,  # change for different metals
             fill = Sex)) +
  geom_bar(stat = "identity",
           position = position_dodge()) +
  geom_errorbar(aes(ymin = Mn - se.lbxbmn, # change for different metalas
                    ymax = Mn + se.lbxbmn)) +
  labs(title="Levels of Blood Mn in Males vs Females Across Life Stages", # edit labels for diff graphs
       x = "Life Stage",
       y = "Average Blood Pb (ug/dL)",
       caption = "NHANES 2015/2016 data")

# Mercury
# mercury data manipulation
hg_data <-
  life_means_data %>%
  mutate("total" = `Total Hg`,
         "inorganic" = `Inorganic Hg`,
         "organic" = `Organic Hg`) %>%
  select(-Pb, -Cd, -Mn, -se.lbxbpb, -se.lbxbcd, -se.lbxbmn) %>%
  pivot_longer(total:organic,
               names_to = "type",
               values_to = "value")

# plot
hg_plot <-
  hg_data %>%
  na.omit() %>%
  ggplot(aes(x = Sex,
             y = value,
             fill = type)) +
  geom_col() +
  facet_grid(. ~life_stage) +
  labs(title="Levels of Blood Hg in Males vs Females Across Life Stages", # edit labels for diff graphs
       y = "Average Blood Hg (ug/dL)",
       caption = "NHANES 2015/2016 data") +
  scale_fill_discrete(name="Mercury Type")
```

list(Sex = c("Male", "Female", "Male", "Female", "Male", "Female", "Male", "Female"), life_stage = c(1, 1, 2, 2, 3, 3, 4, 4), Pb = c(0.8501239, 0.7786317, 0.6394596, 0.473822, 1.2766305, 0.9328629, 1.8764311, 1.4773593), Cd = c(0.0986337, 0.1026739, 0.1770584, 0.1813239, 0.4053038, 0.4629328, 0.4208891, 0.4359469), `Total Hg` = c(0.4459588, 0.4388593, 0.74312, 0.4868935, 1.4361154, 1.4192908, 1.3584639, 1.2410247), `Inorganic Hg` = c(0.210607, 0.2089638, 0.2266632, 0.2406871, 0.2314264, 0.2584671, 0.2573128, 0.2533815), `Organic Hg` = c(0.2353518, 0.2298955, 0.5164568, 0.2462064, 1.204689, 1.1608238, 1.1011512, 0.9876432), Mn = c(10.632091, 11.389825, 10.286462, 11.772936, 9.175066, 10.859249, 9.043854, 9.399506), se.lbxbpb = c(0.0591419, 0.0483942, 0.0190509, 0.0277681, 0.0556552, 0.0416805, 0.0755933, 0.0643802), se.lbxbcd = c(0.0023266, 0.0027021, 0.0158446, 0.0180578, 0.0235644, 0.0185526, 0.0299166, 0.0178966), se.lbxthg = c(0.0289284, 0.0318108, 0.1212807, 0.0323583, 0.0998412, 0.1117299, 0.1060895, 0.1686052), se.lbxihg = c("0.0052623", "0.0070838", "0.0131094", "0.0297582", "0.0071605", "0.0105571", "0.0155124", "0.0176039"), se.lbxohg = c(0.0176039, 0.0311024, 0.1207636, 0.0547276, 0.1017978, 0.1156543, 0.1067021, 0.169314), se.lbxbmn = c(0.1958392, 0.2087941, 0.314473, 0.4017955, 0.1596139, 0.1047459, 0.2692876, 0.263964)), list(,), , list(x = ~life_stage, y = ~Pb, fill = ~Sex), list(), , , , list(x = "Life Stage", y = "Average Blood Pb (ug/dL)", title = "Levels of Blood Pb in Males vs Females Across Life Stages", caption = "NHANES 2015/2016 data", fill = "Sex", ymin = "Pb - se.lbxbpb", ymax = "Pb + se.lbxbpb")

list(Sex = c(“Male”, “Female”, “Male”, “Female”, “Male”, “Female”, “Male”, “Female”), life_stage = c(1, 1, 2, 2, 3, 3, 4, 4), Pb = c(0.8501239, 0.7786317, 0.6394596, 0.473822, 1.2766305, 0.9328629, 1.8764311, 1.4773593), Cd = c(0.0986337, 0.1026739, 0.1770584, 0.1813239, 0.4053038, 0.4629328, 0.4208891, 0.4359469), `Total Hg` = c(0.4459588, 0.4388593, 0.74312, 0.4868935, 1.4361154, 1.4192908, 1.3584639, 1.2410247), `Inorganic Hg` = c(0.210607, 0.2089638, 0.2266632, 0.2406871, 0.2314264, 0.2584671, 0.2573128, 0.2533815), `Organic Hg` = c(0.2353518, 0.2298955, 0.5164568, 0.2462064, 1.204689, 1.1608238, 1.1011512, 0.9876432), Mn = c(10.632091, 11.389825, 10.286462, 11.772936, 9.175066, 10.859249, 9.043854, 9.399506), se.lbxbpb = c(0.0591419, 0.0483942, 0.0190509, 0.0277681, 0.0556552, 0.0416805, 0.0755933, 0.0643802), se.lbxbcd = c(0.0023266, 0.0027021, 0.0158446, 0.0180578, 0.0235644, 0.0185526, 0.0299166, 0.0178966), se.lbxthg = c(0.0289284, 0.0318108, 0.1212807, 0.0323583, 0.0998412, 0.1117299, 0.1060895, 0.1686052), se.lbxihg = c(“0.0052623”, “0.0070838”, “0.0131094”, “0.0297582”, “0.0071605”, “0.0105571”, “0.0155124”, “0.0176039”), se.lbxohg = c(0.0176039, 0.0311024, 0.1207636, 0.0547276, 0.1017978, 0.1156543, 0.1067021, 0.169314), se.lbxbmn = c(0.1958392, 0.2087941, 0.314473, 0.4017955, 0.1596139, 0.1047459, 0.2692876, 0.263964)), list(,), , list(x = ~life_stage, y = ~Cd, fill = ~Sex), list(), , , , list(x = “Life Stage”, y = “Average Blood Pb (ug/L)”, title = “Levels of Blood Cd in Males vs Females Across Life Stages”, caption = “NHANES 2015/2016 data”, fill = “Sex”, ymin = “Cd - se.lbxbcd”, ymax = “Cd + se.lbxbcd”)

list(Sex = c(“Male”, “Female”, “Male”, “Female”, “Male”, “Female”, “Male”, “Female”), life_stage = c(1, 1, 2, 2, 3, 3, 4, 4), Pb = c(0.8501239, 0.7786317, 0.6394596, 0.473822, 1.2766305, 0.9328629, 1.8764311, 1.4773593), Cd = c(0.0986337, 0.1026739, 0.1770584, 0.1813239, 0.4053038, 0.4629328, 0.4208891, 0.4359469), `Total Hg` = c(0.4459588, 0.4388593, 0.74312, 0.4868935, 1.4361154, 1.4192908, 1.3584639, 1.2410247), `Inorganic Hg` = c(0.210607, 0.2089638, 0.2266632, 0.2406871, 0.2314264, 0.2584671, 0.2573128, 0.2533815), `Organic Hg` = c(0.2353518, 0.2298955, 0.5164568, 0.2462064, 1.204689, 1.1608238, 1.1011512, 0.9876432), Mn = c(10.632091, 11.389825, 10.286462, 11.772936, 9.175066, 10.859249, 9.043854, 9.399506), se.lbxbpb = c(0.0591419, 0.0483942, 0.0190509, 0.0277681, 0.0556552, 0.0416805, 0.0755933, 0.0643802), se.lbxbcd = c(0.0023266, 0.0027021, 0.0158446, 0.0180578, 0.0235644, 0.0185526, 0.0299166, 0.0178966), se.lbxthg = c(0.0289284, 0.0318108, 0.1212807, 0.0323583, 0.0998412, 0.1117299, 0.1060895, 0.1686052), se.lbxihg = c(“0.0052623”, “0.0070838”, “0.0131094”, “0.0297582”, “0.0071605”, “0.0105571”, “0.0155124”, “0.0176039”), se.lbxohg = c(0.0176039, 0.0311024, 0.1207636, 0.0547276, 0.1017978, 0.1156543, 0.1067021, 0.169314), se.lbxbmn = c(0.1958392, 0.2087941, 0.314473, 0.4017955, 0.1596139, 0.1047459, 0.2692876, 0.263964)), list(,), , list(x = ~life_stage, y = ~Mn, fill = ~Sex), list(), , , , list(x = “Life Stage”, y = “Average Blood Pb (ug/dL)”, title = “Levels of Blood Mn in Males vs Females Across Life Stages”, caption = “NHANES 2015/2016 data”, fill = “Sex”, ymin = “Mn - se.lbxbmn”, ymax = “Mn + se.lbxbmn”)

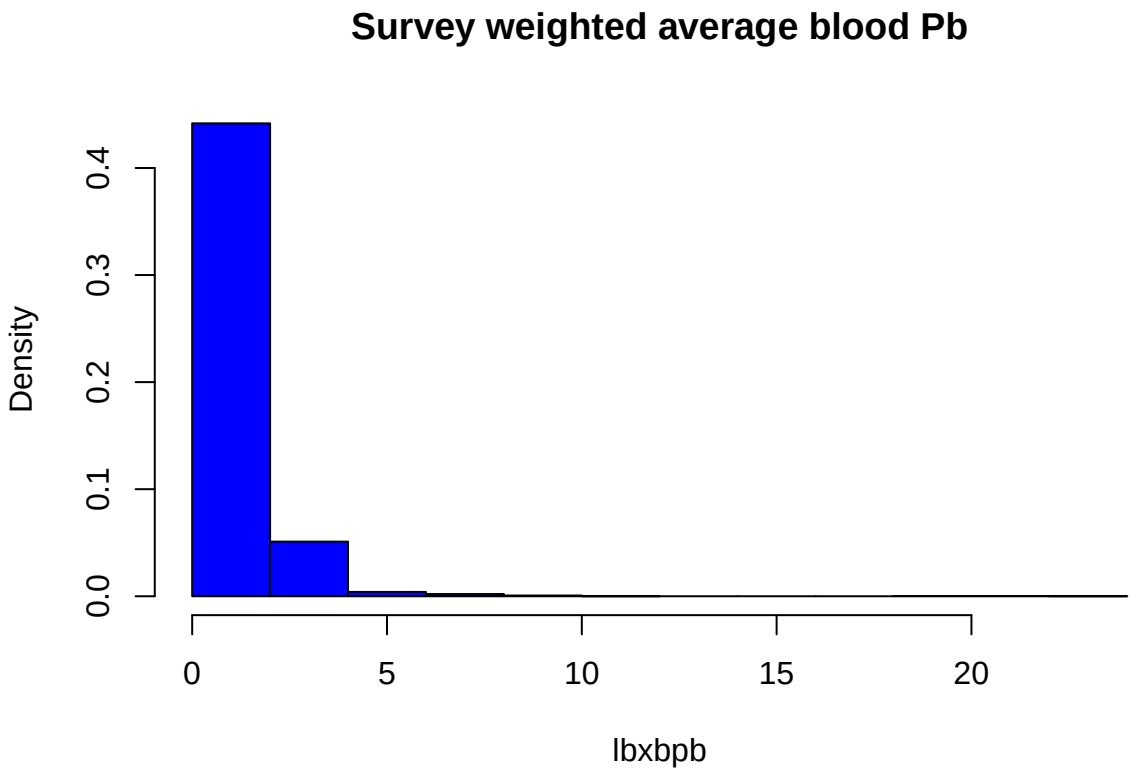
list(Sex = c(“Male”, “Male”, “Male”, “Female”, “Female”, “Female”, “Male”, “Male”, “Male”, “Female”, “Female”, “Female”, “Male”, “Male”, “Male”, “Female”, “Female”, “Female”, “Male”, “Male”, “Male”, “Female”, “Female”, “Female”), life_stage = c(1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4), `Total Hg` = c(0.4459588, 0.4459588, 0.4459588, 0.4388593, 0.4388593, 0.4388593, 0.74312, 0.74312, 0.74312, 0.4868935, 0.4868935, 0.4868935, 1.4361154, 1.4361154, 1.4361154, 1.4192908, 1.4192908, 1.4192908, 1.3584639, 1.3584639, 1.3584639, 1.2410247, 1.2410247, 1.2410247), `Inorganic Hg` = c(0.210607, 0.210607, 0.210607, 0.2089638, 0.2089638, 0.2089638, 0.2266632, 0.2266632, 0.2266632, 0.2406871, 0.2406871, 0.2406871, 0.2314264, 0.2314264, 0.2314264, 0.2584671, 0.2584671, 0.2584671, 0.2573128, 0.2573128, 0.2573128, 0.2533815, 0.2533815, 0.2533815), `Organic Hg` = c(0.2353518, 0.2353518, 0.2353518, 0.2298955, 0.2298955, 0.2298955, 0.5164568, 0.5164568, 0.5164568, 0.2462064, 0.2462064, 0.2462064, 1.204689, 1.204689, 1.204689, 1.1608238, 1.1608238, 1.1608238, 1.1011512, 1.1011512, 1.1011512, 0.9876432, 0.9876432, 0.9876432), se.lbxthg = c(0.0289284, 0.0289284, 0.0289284, 0.0318108, 0.0318108, 0.0318108, 0.1212807, 0.1212807, 0.1212807, 0.0323583, 0.0323583, 0.0323583, 0.0998412, 0.0998412, 0.0998412, 0.1117299, 0.1117299, 0.1117299, 0.1060895, 0.1060895, 0.1060895, 0.1686052, 0.1686052, 0.1686052), se.lbxihg = c(“0.0052623”, “0.0052623”, “0.0052623”, “0.0070838”, “0.0070838”, “0.0070838”, “0.0131094”, “0.0131094”, “0.0131094”, “0.0297582”, “0.0297582”, “0.0297582”, “0.0071605”, “0.0071605”, “0.0071605”, “0.0105571”, “0.0105571”, “0.0105571”, “0.0155124”, “0.0155124”, “0.0155124”, “0.0176039”, “0.0176039”, “0.0176039”), se.lbxohg = c(0.0176039, 0.0176039, 0.0176039, 0.0311024, 0.0311024, 0.0311024, 0.1207636, 0.1207636, 0.1207636, 0.0547276, 0.0547276, 0.0547276, 0.1017978, 0.1017978, 0.1017978, 0.1156543, 0.1156543, 0.1156543, 0.1067021, 0.1067021, 0.1067021, 0.169314, 0.169314, 0.169314), type = c(“total”, “inorganic”, “organic”, “total”, “inorganic”, “organic”, “total”, “inorganic”, “organic”, “total”, “inorganic”, “organic”, “total”, “inorganic”, “organic”, “total”, “inorganic”,

"organic", "total", "inorganic", "organic", "total", "inorganic", "organic"), value = c(0.4459588, 0.210607, 0.2353518, 0.4388593, 0.2089638, 0.2298955, 0.74312, 0.2266632, 0.5164568, 0.4868935, 0.2406871, 0.2462064, 1.4361154, 0.2314264, 1.204689, 1.4192908, 0.2584671, 1.1608238, 1.3584639, 0.2573128, 1.1011512, 1.2410247, 0.2533815, 0.9876432)), list(), , list(x = ~Sex, y = ~value, fill = ~type), list(), , , , list(y = "Average Blood Hg (ug/dL)", title = "Levels of Blood Hg in Males vs Females Across Life Stages", caption = "NHANES 2015/2016 data", x = "Sex", fill = "type")

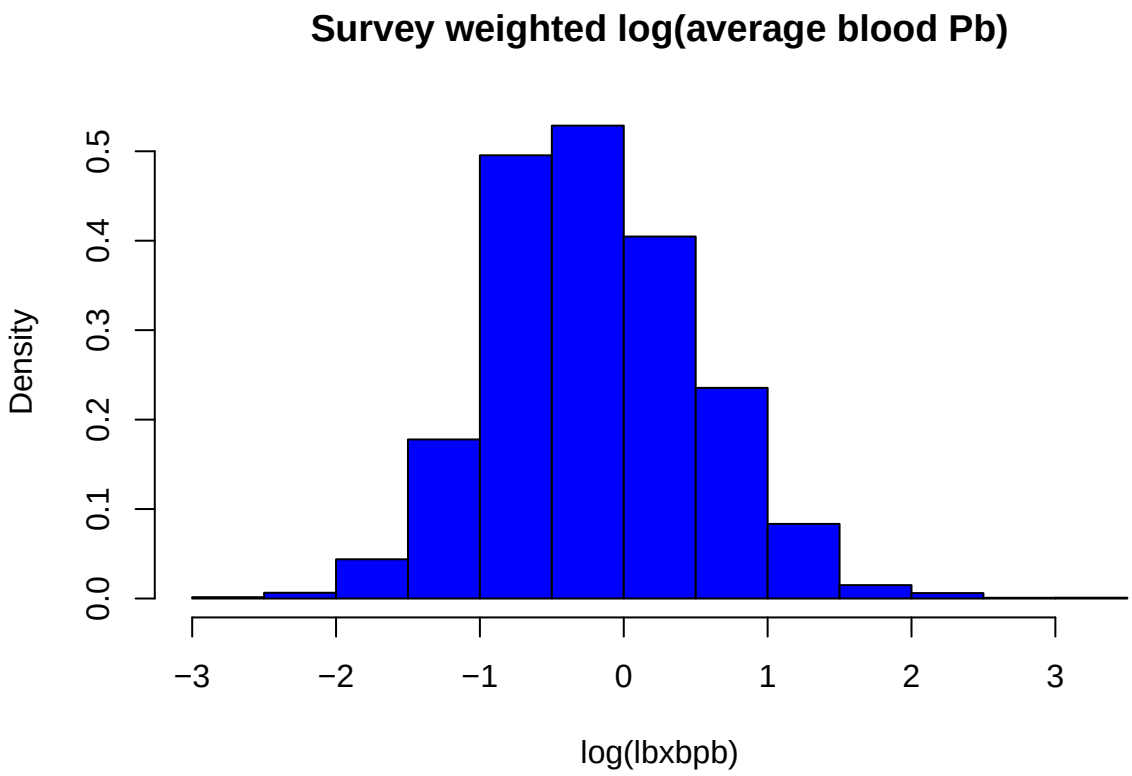
```
# is

# We are comparing a binary variable (hypertension status) across a categorical variable, so let's use
nhanes.svy <-
  svydesign(ids = ~sdmvpsu,
            strata = ~sdmvstra,
            weights = ~wtsh2yr,
            nest = TRUE,
            data = nh1516_sub) #Re-setting the svydesign!

# Contingency table and chi-squared tests for survey data using svychisq
table(nh1516_sub$htn,
      nh1516_sub$race)

svychisq(~htn+sex,
         nhanes.svy,
         statistic="adjWald")
```

Data Visualization

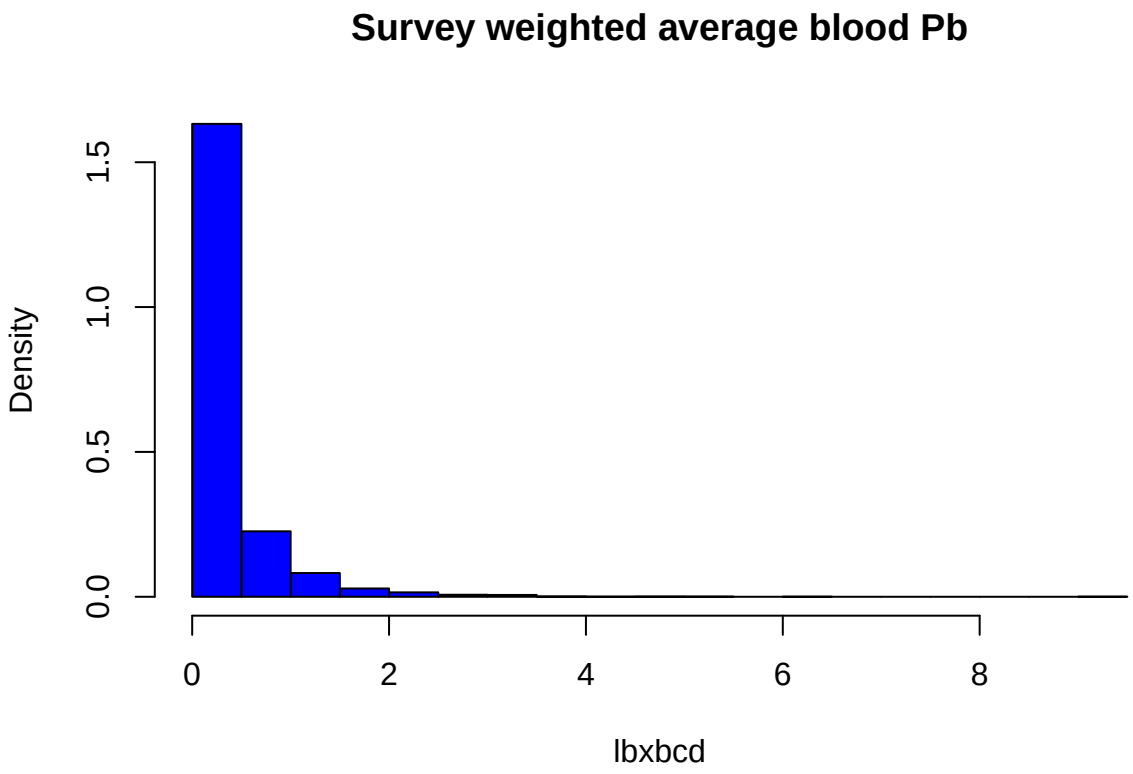
```
# Lead
##############################################

# "svyhist" creates histograms accounting for survey weights
svyhist(~lbxbpb,
        nhanes.svy,
        main="Survey weighted average blood Pb",col="blue")
```



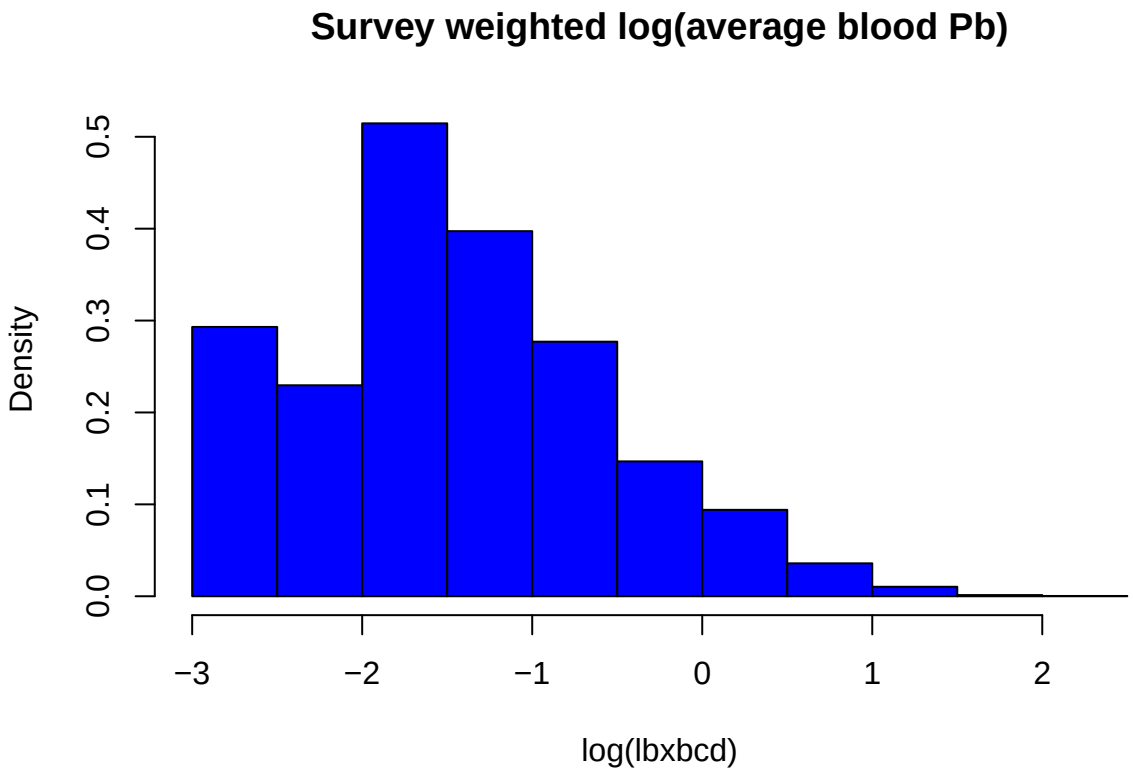

```
# log transformed
svyhist(~log(lbxbpb),
        nhanes.svy,
        main="Survey weighted log(average blood Pb)",col="blue")
```



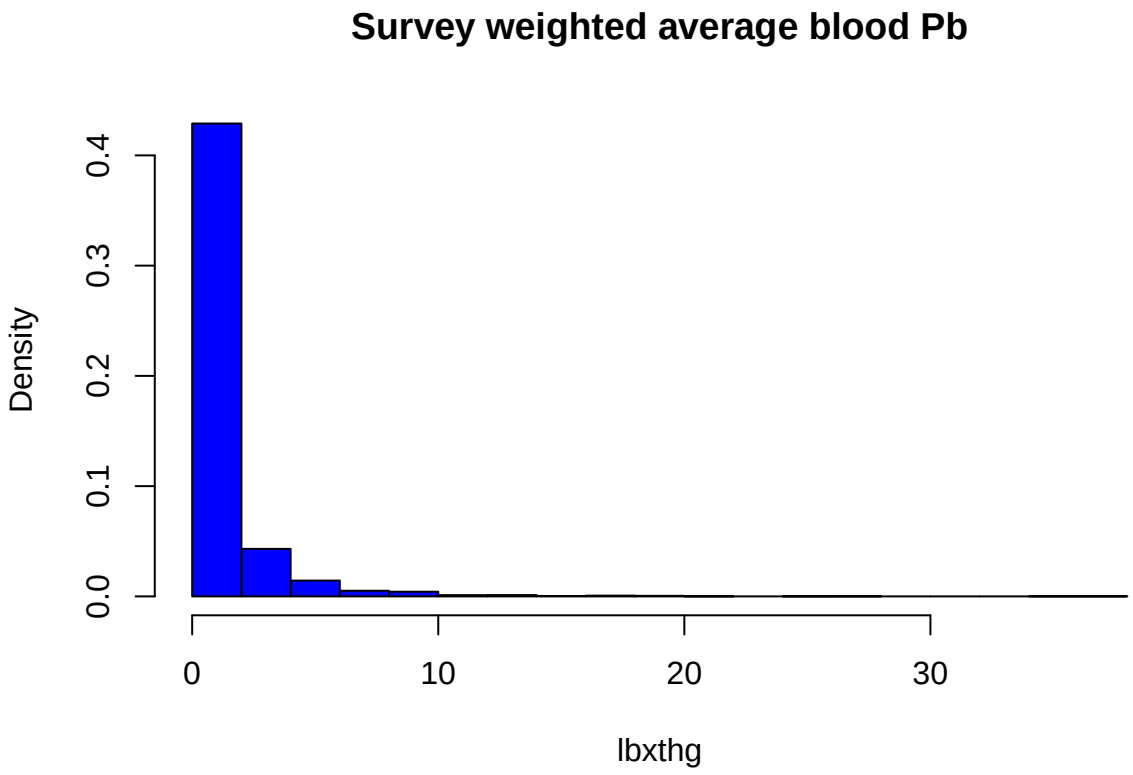
```r
# Cadmium
##########################################
svyhist(~lbxbcd,
        nhanes.svy,
        main="Survey weighted average blood Pb",col="blue")
```



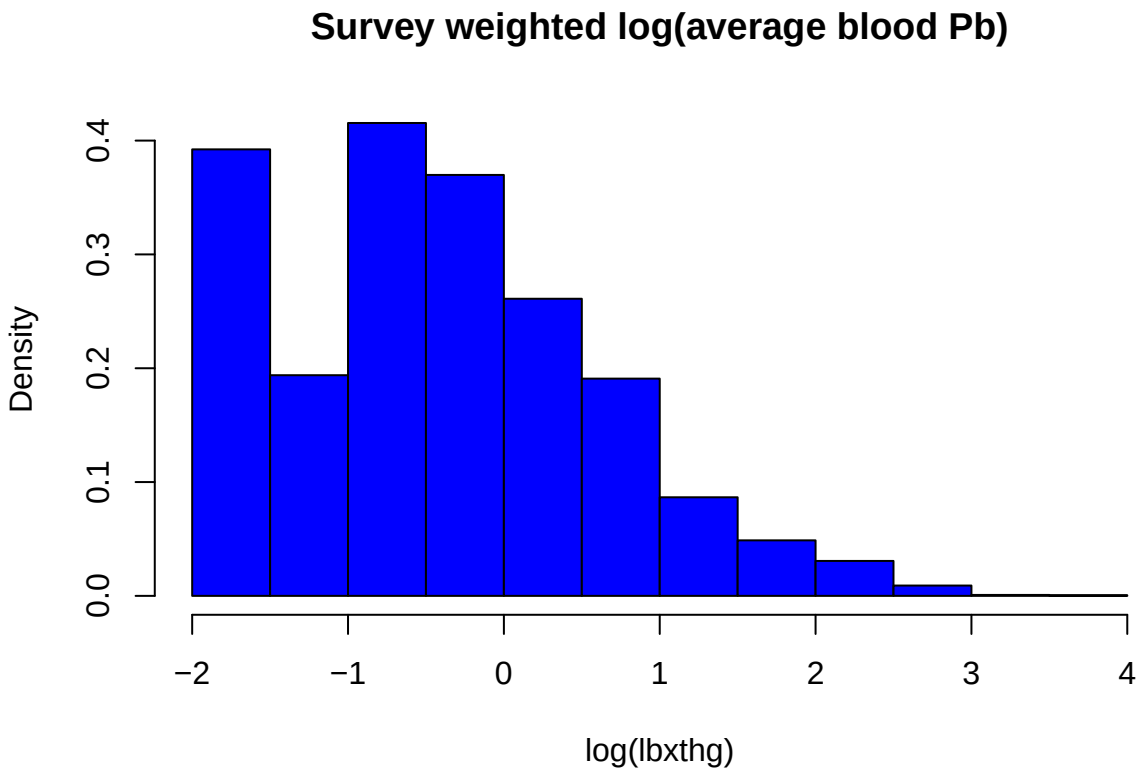

```r
# log transformed
svyhist(~log(lbxbcd),
        nhanes.svy,
        main="Survey weighted log(average blood Pb)",col="blue")
```



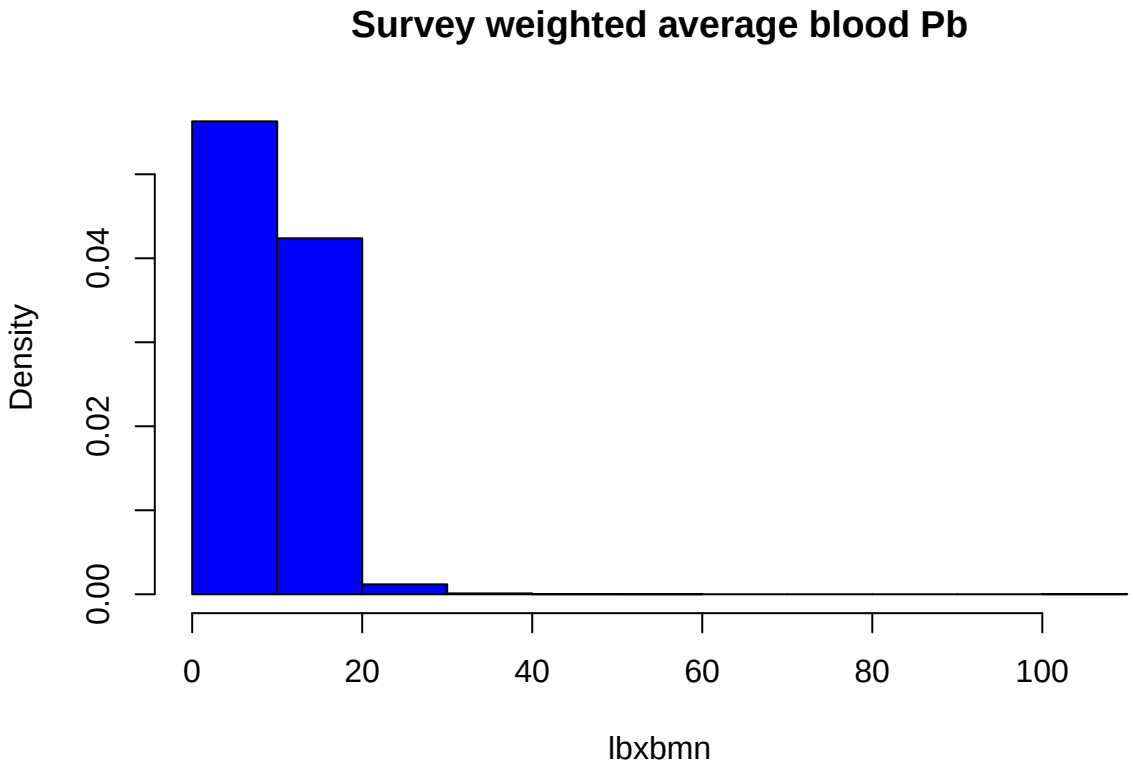

```
# Mercury
##############################################
svyhist(~lbxthg,
        nhanes.svy,
        main="Survey weighted average blood Pb",col="blue")
```



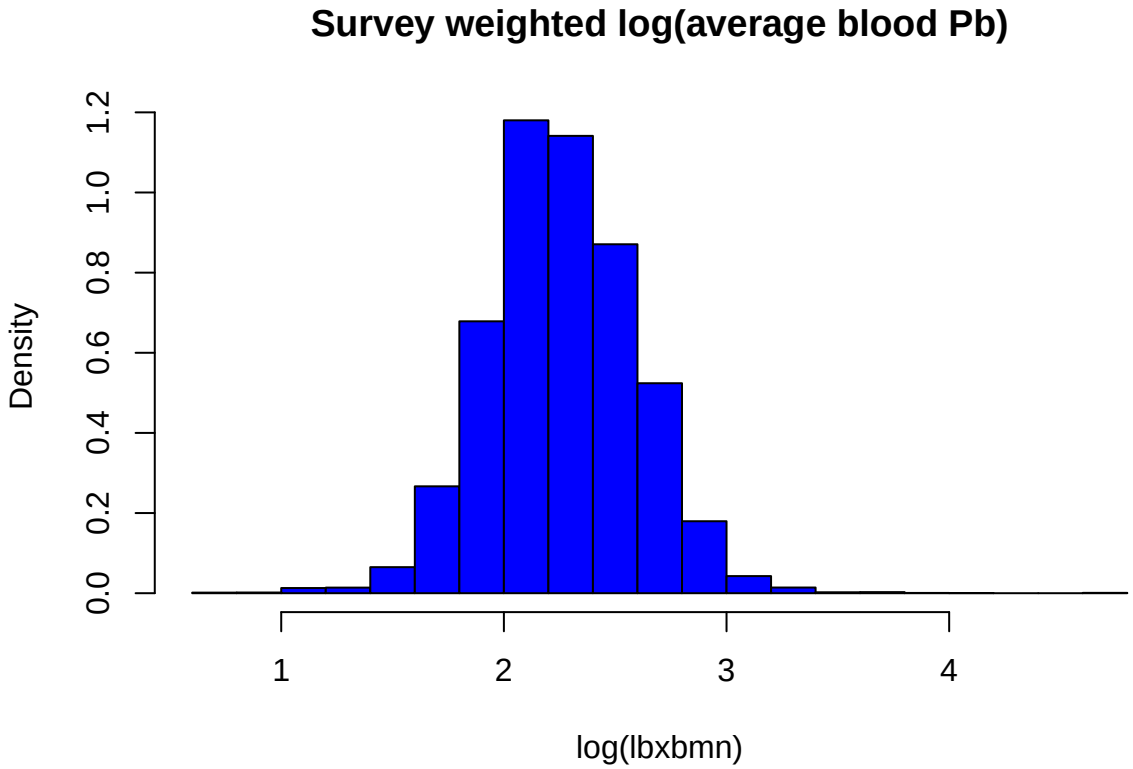
```r
# log transformed
svyhist(~log(lbxthg),
        nhanes.svy,
        main="Survey weighted log(average blood Pb)",col="blue")
```




```r
# Manganese
##############################################
svyhist(~lbxbmn,
        nhanes.svy,
        main="Survey weighted average blood Pb",col="blue")
```




```
# log transformed
svyhist(~log(lbxbmn),
        nhanes.svy,
        main="Survey weighted log(average blood Pb)",col="blue")
```



```
# log transformed distributions look more normal!
```